Parsons *et al. BMC Res Notes (2017) 10:753*
https://doi.org/10.1186/s13104-017-3089-4

BMC Research Notes

RESEARCH NOTE

Open Access

Identification of 24 new microsatellite loci in the sweat bee *Lasioglossum malachurum* (Hymenoptera: Halictidae)

Paul J. Parsons[1,2*], Christelle Couchoux[1], Gavin J. Horsburgh[2], Deborah A. Dawson[2] and Jeremy Field[1]

Abstract

Objective: The objective here is to identify highly polymorphic microsatellite loci for the Palaearctic sweat bee *Lasioglossum malachurum*. Sweat bees (Hymenoptera: Halictidae) are widespread pollinators that exhibit an unusually large range of social behaviours from non-social, where each female nests alone, to eusocial, where a single queen reproduces while the other members of the colony help to rear her offspring. They thus represent excellent models for understanding social evolution.

Results: 24 new microsatellite loci were successfully optimized. When amplified across 23–40 unrelated females, the number of alleles per locus ranged from 3 to 17 and the observed heterozygosities 0.45 to 0.95. Only one locus showed evidence of significant deviation from Hardy–Weinberg equilibrium. No evidence of linkage disequilibrium was found. These 24 loci will enable researchers to gain greater understanding of colony relationships within this species, an important model for the study of eusociality. Furthermore, 22 of the same loci were also successfully amplified in *L. calceatum*, suggesting that these loci may be useful for investigating the ecology and evolution of sweat bees in general.

Keywords: Halictidae, Microsatellite, *Lasioglossum malachurum*, *Lasioglossum calceatum*, Sweat bee

Introduction

Sweat bees (Hymenoptera: Halictidae) are widespread pollinators which exhibit an unusually large range of social behaviours from non-social, where each female nests alone, to eusocial, where a single queen reproduces while the other members of the colony help to rear her offspring [1]. Sweat bees are also unusual in that social and non-social species are often closely related, with multiple evolutionary transitions having occurred between sociality and non-sociality [2]. Sweat bees thus represent excellent models for understanding social evolution [1, 2]. Here we present a new set of microsatellite loci developed from *Lasioglossum malachurum* (Kirby, 1802), a haplodiploid eusocial species that has been particularly well studied, mainly because it is widely distributed in the Western Palaearctic and because it often occurs in large, dense nesting aggregations that facilitate behavioural research [3–6]. Microsatellite markers are widely used in social evolution research, for example to investigate population structure, estimate genetic relatedness and assign offspring to parents [7–9]. Microsatellite loci have been developed for this species previously [3, 10] but most of them have comparatively low heterozygosities and are difficult to combine into multiplex reactions because of highly specific annealing temperatures and polymerase chain reaction (PCR) mixes. Here, we report 24 new microsatellite markers developed for *L. malachurum*, 14 of which have been efficiently amplified in two multiplex sets. These markers should substantially aid future studies on sweat bee behaviour and ecology.

Main text

Lasioglossum malachurum females were sampled from a field site at Denton in East Sussex, UK in 2015. Genomic DNA was extracted from head, abdomen

*Correspondence: p.parsons@exeter.ac.uk
[1] Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Cornwall TR10 9EZ, UK
Full list of author information is available at the end of the article

and/or legs using an ammonium acetate extraction method [11, 12]. DNA concentration was quantified using a Fluostar Optima fluorimeter and its quality assessed using gel electrophoresis. DNA from one foundress (female M4) from Denton was digested using *MboI* and the fragments enriched for dinucleotide and tetranucleotide repeat motifs (following [13]). An Illumina paired end library was then compiled using this repeat-enriched genomic DNA. The NEBNext Ultra library preparation kit (New England Biolabs Inc. Cat. No. E7370S) protocol was followed and DNA sequencing was conducted using a MiSeq Benchtop Sequencer (Illumina). Primer sets were designed from 53 microsatellite sequences using PRIMER3 v0.4.0 [14]. Sequences were confirmed to be unique using BLAST software [15].

Each 2 µl PCR contained approximately 10 ng of air-dried genomic DNA, 0.2 µM of each primer and 1 µl QIAGEN Multiplex PCR mix (QIAGEN Inc. Cat. No. 20614) following [16]. As we required loci that could be reliably multiplexed together for efficient use we designed primers with very similar melting temperatures (± 2 °C) enabling these to be amplified at the same annealing temperature (57 °C). The following PCR profile was used: 95 °C for 15 min, followed by 44 cycles of 94 °C for 30 s, 57 °C for 90 s, 72 °C for 90 s and finally 60 °C for 30 min. PCR amplification was performed using a DNA Engine Tetrad ®Thermal Cycler (MJ Research, Bio-Rad, Hemel Hempstead, Herts, UK). PCR products were genotyped on an ABI 3730 48-well capillary DNA Analyser using the LIZ size standard (Applied Biosystems Inc. Cat. No. 4322682). Alleles were scored using GENEMAPPERv3.7 software (Applied Biosystems Inc.). Of the 53 markers, 24 could be scored reliably across the test sample (23–40 females all from the same field site at Denton) (Table 1). The remaining 29 were found to be either monomorphic or unreliable following our PCR methodology (Table 2). It is possible that with more specific optimization, some of these could be used in future studies. We successfully incorporated 14 of the optimized markers into two multiplex panels (using the above PCR reagents and concentrations) with no dropout or artifacts produced (Table 1).

The numbers of alleles and heterozygosities were calculated for each of the 24 loci using CERVUS v3.0.6 and with the sample sizes shown in Table 1 [17]. Tests for deviation from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were conducted using GENEPOP web version 4.2 [18]. To correct p-values in multiple tests, the Q Value was applied to LD p-values. The q value is a measure of the significance in terms of false discovery rate, rather than conventional Bonferroni correction which attempts to measure significance in terms of false positives only [19]. Observed levels of heterozygosity ranged from 0.45 to 0.95 with 3–17 alleles per locus (Table 1). Only *Lma31* deviated from HWE ($p = 0.049$). No groups of loci displayed LD, providing no evidence of physical linkage based on the individuals genotyped.

These loci are likely to be useful for investigating the ecology and behaviour of *L. malachurum* and also potentially that of other sweat bees. Indeed, we have successfully amplified 22 of the 24 loci in *L. calceatum* (Scopoli) individuals sampled in the UK; only *Lma20* and *Lma21* failed to amplify and 17 of the 22 loci that did amplify were polymorphic (Table 1; Davison & Field, in prep.).

Table 1 Characterisation of 24 new *L. malachurum* microsatellites

Locus name	GenBank sequence accession number	Panel and dye	Repeat motif	Primer sequence (5′–3′)	N[a] tested	N alleles	Expected allele size[b], size range	HObs	HExp	HWE *p* value	Est. F (null)	L. cal success[c]
Lma02	MG273262	1 NED	$(TC)_{13}$	F: CCGAGTTCATCAACATCCTC R: TTGATTATCAGCGAGATGAGC	23	10	150 139–185	0.87	0.83	0.712	− 0.037	P
Lma03	MG273263	1 PET	$(AG)_{14}$	F: AAAGCGTTGCGAGACACC R: AGCATAATGGAAACCCAACG	38	7	154 137–167	0.816	0.745	0.103	− 0.063	P
Lma04	MG273264		$(TG)_{12}$	F: CGTTACCGCGTTGGTTTC R: GTCTTGTCTAACCGCAACAGC	37	6	169 165–177	0.649	0.727	0.162	0.034	M
Lma12	MG273265	2 PET	$(CT)_{12}$	F: CCAACCGAACACCAACTTTC R: CTCCCGGGTTGTCATGTAAG	39	10	150 131–181	0.667	0.701	0.413	0.017	P
Lma14	MG273266	1 6-FAM	$(AG)_{14}$	F: CAACGCGTGACAGGTGATAC R: CGGCTACGTTCCACTATGAAG	40	14	170 162–192	0.825	0.896	0.186	0.035	P
Lma20	MG273267		$(AG)_{19}$	F: AGCGCTCGATGACTGTCG R: TTGCGCAAGCCGTTCTAC	39	17	210 196–262	0.872	0.889	0.087	0.007	F
Lma21	MG273268	2 NED	$(GA)_{16}$	F: CGGTAAACTTGCTTCGACCTG R: CCGATTCCTTCACAGACACG	38	11	137 135–156	0.868	0.85	0.053	− 0.026	F
Lma23	MG273269		$(GA)_{13}$	F: GATAATCAATGGTAATCGGTTGG R: TTAACATCGTTCGCTTCTCG	40	11	167 154–218	0.85	0.838	0.179	− 0.017	M
Lma24	MG273270	2 6-FAM	$(GA)_{13}$ CA $(GA)_6$	F: TCCTCGGACAAGGAGATACG R: TTCGGGTACCGTTCAGTCTC	40	13	172 141–181	0.925	0.891	0.723	− 0.026	P
Lma27	MG273271		$(GA)_{13}$	F: GCTGGCAGCTCTGGAGAAG R: TGACGGCCATTTAGTTCGTC	38	9	189 177–199	0.737	0.804	0.071	0.032	P
Lma29	MG273272	1 PET	$(CT)_4$ TT $(CT)_9$	F: CTCGTCCCTCGTGTGACTC R: GTATCGTGCGTGCGTGTC	38	12	204 201–231	0.868	0.883	0.725	0.003	P
Lma30	MG273273		$(GACGA)_6$	F: TCCGTCTCTGGTCGATACTG R: ACAGCAGCATCTGAACTTGC	38	3	237 225–235	0.447	0.407	0.854	− 0.075	P
Lma31	MG273274		$(TCTT)_{10}$	F: CGCACTCCGCTTTTCCTC R: CGTCACCAGGAGAGCAAGG	40	6	146 142–164	0.55	0.664	0.049	0.084	P
Lma34	MG273275		$(CT)_{12}$	F: TCTGAACAGTACGGAACAATGC R: ACCGACACGGGAGAGAGAG	40	6	176 165–179	0.675	0.684	0.718	− 0.009	P
Lma36	MG273285	1 VIC	$(CT)_{16}$	F: GGCCCTTCGACTTTGTTG R: GAATCTCTGGGTGCTCTAACG	38	8	188 185–199	0.737	0.785	0.298	0.027	P
Lma39	MG273276	2 PET	$(CTAT)_8$	F: CGAGCCTATGCAGAGAACAG R: TGGATGGCTGCTGAGTAAAC	38	7	205 205–237	0.789	0.75	0.68	− 0.034	P

Table 1 continued

Locus name	GenBank sequence accession number	Panel and dye	Repeat motif	Primer sequence (5′–3′)	N[a] tested	N alleles	Expected allele size[b], size range	HObs	HExp	HWE *p* value	Est. F (null)	L. cal success[c]
Lma40	MG273277	2	$(GA)_{12}$	F: CGTTCGTTCGTTCGTTACTG	38	14	150	0.947	0.906	0.74	− 0.029	P
		VIC		R: CAGAGTGCGTCGCTTGTTAG			155–189					
Lma42	MG273278		$(AG)_{13}$	F: ACCATCGCCCTTCCACTAC	40	5	167	0.75	0.733	0.623	− 0.016	P
				R: CCGAAACTATTCGCCCATC			161–169					
Lma48	MG273279	2	$(TC)_{14}$	F: GTTGGATGCATCTGGAGGAC	38	6	206	0.763	0.722	0.14	− 0.043	M
		NED		R: TGCGGTGGTTATTGATTTCC			193–209					
Lma49	MG273280		$(GAAA)_{10}$	F: GAGAGGGTGGTTGCACTACG	38	4	209	0.684	0.62	0.786	− 0.055	M
				R: CTCGTGGAATCGAACTCACC			189–209					
Lma50	MG273281		$(CT)_3$ CG $(CT)_{12}$	F: CGTTTAACCGGCTCGCTAC	38	8	181	0.684	0.763	0.726	0.051	P
				R: CCGCGAATAAGTGGAGTGTC			163–209					
Lma51	MG273282	1	$(CT)_{11}$	F: GAGAAATTGCCAGCAAACATC	40	4	243	0.475	0.545	0.259	0.066	P
		6-FAM		R: AGTTTCGTGGAAGGGAACG			237–243					
Lma52	MG273283	1	$(TG)_{11}$	F: CGGCAACTGCTTGCATAAC	40	5	156	0.8	0.732	0.575	− 0.056	M
		VIC		R: CCCGTAGCACTCGCATACTC			151–159					
Lma53	MG273284	1	$(AC)_{12}$	F: ACGCGGGATTACTTTCAATC	40	9	228	0.675	0.759	0.053	0.057	P
		NED		R: CCAATTATCGGGTGAAGGAG			217–241					

[a] N: number of diploid, unrelated *L. malachurum* females genotyped (all from the same population at Denton)

[b] Based on the sequenced individual (sample M4); Hobs and HExp: observed and expected heterozygosities; HWE: *p* value when testing for deviation from Hardy–Weinberg equilibrium; F(Null): Estimated frequency of null alleles

[c] Amplification success across 14 *L. calceatum* individuals: *F* failed to amplify, *M* monomorphic, *p* polymorphic

Table 2 Identification of a further 29 markers that were rejected and not considered for multiplex panels

Locus name	GenBank sequence accession number	Repeat motif	Primer sequence	Expected allele size[†]	Reason for dropping (tested in 23–24 individuals)
Lma01	MG273287	$(TGAC)_7$	F: AACGCCTCGGTGAACCTG R: TCGAGTTCTCCCTCCTCGTATC	108	Monomorphic
Lma05	MG273288	$(TTTC)_7$	F: ATGCGTCTAAATCGTTCCTG R: AACAAAGAATGAACGAACGTG	178	Monomorphic
Lma06	MG273289	$(AG)_{11}$	F: CGGGAACGACGGAGAGAG R: ACGGGTCTGTTCACCCTTTG	184	False peaks
Lma07	MG273286	$(GAAA)_5$	F: GTCATGGAGAGGGTGGTTG R: CAATCTCAACCGTGTTCGTC	189	No product
Lma08	MG273290	$(TTCT)_7$	F: CTATCCGAGGCCTGTACACTG R: ATCTGAAATCGTGGCTGGTC	192	No product
Lma09	MG273291	$(AGAA)_5$	F: ACGGGACTGAAAGGGACAC R: TACTTCGCGTGCCTGTCTC	201	Monomorphic
Lma10	MG273292	$(AAAG)_7$	F: GAGACAACGAGGGAGAAAGC R: AACCTCAACCGTGTTCGTC	206	Stutter
Lma11	MG273293	$(GA)_7$	F: CTTGTACCACGCGTACACACC R: GCCCTGCGTCTTCTCCTC	111	False peaks
Lma13	MG273294	$(CT)_{18}$	F: GCTCATCGAGGACGAGGTG R: GCGGTTGGCTGTCATAAGTG	154	False peaks
Lma15	MG273295	$(TGCT)_5$	F: GGACAGTCCGACGAAGGAG R: GCTTCATCCCTTTACTCCATAGC	179	No product
Lma16	MG273296	$(TC)_{20}$	F: ACATTGTTCACCGGACAAATC R: CGTCGAGGATAAGGTTACGG	187	Monomorphic
Lma17	MG273297	$(TC)_{11}$	F: GTCAACGGTAATCCGAGGTG R: TGATACACCGGGAACCATTC	189	False peaks/Stutter
Lma18	MG273298	$(AG)_{16}$	F: GGGATACTAGACAGCCGGAATATAG R: GAATGAACCACGCCGAAG	193	False peaks
Lma19	MG273299	$(TC)_{20}$	F: TGTAAACGGCCGAAGTGTC R: ACAATGTGTGTTCCGGTCAG	203	False peaks/Stutter
Lma22	MG273300	$(TCTT)_5$	F: GCCGGACCAGATTAAATGC R: AAAGACGAGGCTCAAAGAAGC	151	No product
Lma25	MG273301	$(CTAT)_6$	F: CGAAATACCGTTAACCAACATC R: TAAAGTGGCGAGTGATGGAC	180	Monomorphic
Lma26	MG273302	$(AG)_{16}$	F: CTTCGATTCCTCGGGTCAC R: TTCCGGCACGTTTATGTAGC	188	No product
Lma28	MG273303	$(CT)_{17}$	F: ATTCGCGACAATGAACGAG R: CAAACGCGAGTCAATAAATCC	193	False peaks

Table 2 continued

Locus name	GenBank sequence accession number	Repeat motif	Primer sequence	Expected allele size[†]	Reason for dropping (tested in 23–24 individuals)
Lma32	MG273304	$(TC)_{16}$	F: CGACGTACCTCTGCTTCCTC R: AGGTCACTTAAATGGTGGTTGG	152	Stutter
Lma33	MG273305	$(GA)_{19}$	F: CTCTTCTCGATTCCGTCTGG R: TTTCGGCTCTTTGCTCTCTC	167	False peaks
Lma35	MG273306	$(GAGT)_5$	F: CCTTCGAGAGGTCAGAGCTAAAG R: CACGTGGCACCACAAATTC	181	No product
Lma37	MG273307	$(TTCT)_5$	F: GTGGCCTATGCTCCTCTCC R: ATCTGAAATCGTGGCTGGTC	190	Monomorphic
Lma38	MG273308	$(GACA)_9$	F: AGAGACAAAGGCGGAGACAG R: TATCTGCGAGACCGACGA	197	False peaks/Stutter
Lma41	MG273309	$(TC)_{20}$	F: AATGATTGTGAACAGTTTGGTATG R: CGAGACTGCAAGAAGTTTCAC	152	Stutter
Lma43	MG273310	$(AG)_{17}$	F: TTCAGCCGAGGGTAGCAC R: CGTACCATCATCTCGTGTCG	178	False peaks
Lma44	MG273311	$(AG)_{15}$	F: ATGAGACTGGCACGACTGTG R: ATGCGTCGCTCCCTTAATC	182	False peaks
Lma45	MG273312	$(CT)_{15}$	F: TTTCGCATCCATCTTCCTTC R: CGCGAATTTCGGTATCTTTC	189	False peaks/Stutter
Lma46	MG273313	$(TCCT)_5$	F: TCCCTTTACCTTCCTTTCTCG R: TGCAACATTTGTACCGAACAG	190	Monomorphic
Lma47	MG273314	$(CTTT)_5$	F: CTATCCGAGGCCTGTACACTG R: GGGTAAGCAAGCATCGTTTC	197	Stutter

[†] Based on the sequenced individual (sample M4)

Limitations

Due to the relatively short read length of the MiSeq Benchtop Sequencing system we were unable to design primer sets to amplify greater than 300 bases. This may however be somewhat fortuitous; the incorporation of larger markers into multiplex panels often proves problematic, since they are generally harder to amplify than markers with smaller products and are more susceptible to dropout [20].

Abbreviations

HWE: Hardy–Weinberg equilibrium; LD: linkage disequilibrium; PCR: polymerase chain reaction.

Authors' contributions

PP: Tested and optimised the primers, scored loci, performed the analysis, co-wrote the paper. CC: Collected samples, scored loci and assisted with the analysis. GJH: prepared the MiSeq library and designed primer sets. DAD: Assisted with marker development, primer design, analysis and manuscript preparation. JF: Wrote the grant application, collected samples and co-wrote the paper. All authors read and approved the final manuscript.

Author details

[1] Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Cornwall TR10 9EZ, UK. [2] NERC Biomolecular Analysis Facility, Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

Acknowledgements

We thank Paul Davison for assisting with sample collection and the NERC Biomolecular Analysis Facility—Sheffield, UK for their facilities.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The sequences acquired herein have been submitted to Genbank, Accession Numbers MG273262–MG273314.

Consent for publication

Not applicable

Ethics approval and consent to participate

Not applicable

Funding

This work was funded by the UK Natural Environment Research Council (NERC Grant NE/M003191/1 to J.F.). This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 695744 to JF).

Publisher's Note

Received: 24 August 2017 Accepted: 13 December 2017
Published online: 19 December 2017

References

1. Schwarz MP, Richards MH, Danforth BN. Changing paradigms in insect social evolution: insights from halictine and allodapine bees. Annu Rev Entomol. 2007;52:127–50.
2. Danforth BN. Evolution of sociality in a primitively eusocial lineage of bees. Proc Natl Acad Sci USA. 2002;99:286–90.
3. Paxton RJ, Arevalo E, Field J. Microsatellite loci for the eusocial *Lasioglossum malachurum* and other sweat bees (Hymenoptera, Halictidae). Mol Ecol Notes. 2003;3:82–4.
4. Strohm E, Bordon-Hauser A. Advantages and disadvantages of large colony size in a halictid bee: the queen's perspective. Behav Ecol. 2003;14:546–53.
5. Wyman LM, Richards MH. Colony social organization of *Lasioglossum malachurum* Kirby (Hymenoptera, Halictidae) in southern Greece. Insect Soc. 2003;50:201–11.
6. Zobel MU, Paxton RJ. Is big the best? Queen size, usurpation and nest closure in a primitively eusocial sweat bee (*Lasioglossum malachurum*). Behav Ecol Sociobiol. 2007;61:435–47.
7. Paxton RJ, Ayasse M, Field J, Soro A. Complex sociogenetic organization and reproductive skew in a primitively eusocial sweat bee, *Lasioglossum malachurum*, as revealed by microsatellites. Mol Ecol. 2002;11:2405–16.
8. Field J, Paxton RJ, Soro A, Bridge C. Cryptic plasticity underlies a major evolutionary transition. Curr Biol. 2010;20:2028–31.
9. Soro A, Field J, Bridge C, Cardinal SC, Paxton RJ. Genetic differentiation across the social transition in a socially polymorphic sweat bee, *Halictus rubicundus*. Mol Ecol. 2010;19:3351–63.
10. Soro A, Ayasse M, Zobel MU, Paxton RJ. Complex sociogenetic organization and the origin of unrelated workers in a eusocial sweat bee, *Lasioglossum malachurum*. Insect Soc. 2009;56:55–63.
11. Nicholls JA, Double MC, Rowell DM, Magrath RD. The evolution of cooperative and pair breeding in thornbills Acanthiza (Pardalotidae). J Avian Biol. 2000;31:165–76.
12. Richardson DS, Jury FL, Blaakmeer K, Komdeur J, Burke T. Parentage assignment and extra-group paternity in a cooperative breeder: the *Seychelles warbler* (*Acrocephalus sechellensis*). Mol Ecol. 2001;10:2263–73.
13. Armour JAL, Neumann R, Gobert S, Jeffreys AJ. Isolation of human simple repeat loci by hybridization selection. Hum Mol Gen. 1994;3:599–605.
14. Rozen S, Skaletsky HJ. PRIMER version 3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. New Jersey: Humana Press; 2000. p. 365–86.
15. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and psi–BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
16. Kenta T, Gratten J, Haigh NS, Hinten GN, Slate J, Butlin RK, et al. Multiplex SNP-SCALE:a cost-effective medium-throughput single nucleotide polymorphism genotyping method. Mol Ecol Res. 2008;8:1230–8.
17. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol. 2007;16:1099–106.
18. Rousset F. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Res. 2008;8:103–10.
19. Storey J. A direct approach to false discovery rates. J R Stat Soc Series B Stat Methodol. 2002;64:479–98.
20. Toouli CD, Turner DR, Grist SA, Morley AA. The effect of cycle number and target size on polymerase chain reaction amplification of polymorphic repetitive sequences. Anal Biochem. 2000;280:324–6.